

Handwritten initials "HOB" in the top left corner.



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/683,235	12/05/2001	Gregory T. Grefenstette	D/A0A34	8303

25453 7590 12/23/2002

PATENT DOCUMENTATION CENTER
XEROX CORPORATION
100 CLINTON AVE., SOUTH, XEROX SQUARE, 20TH FLOOR
ROCHESTER, NY 14644

EXAMINER

WOO, ISAAC M

ART UNIT	PAPER NUMBER
----------	--------------

2172

DATE MAILED: 12/23/2002

Please find below and/or attached an Office communication concerning this application or proceeding.

MB

Office Action Summary

Application No.

09/683,235

Applicant(s)

GREFENSTETTE ET AL.

Examiner

Isaac M Woo

Art Unit

2172

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 05 December 2001.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-20 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-20 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11) ☐ The proposed drawing correction filed on _____ is: a) ☐ approved b) ☐ disapproved by the Examiner.
If approved, corrected drawings are required in reply to this Office action.
- 12) ☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

- 13) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
* See the attached detailed Office action for a list of the certified copies not received.
- 14) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
a) ☐ The translation of the foreign language provisional application has been received.
- 15) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892) 4) ☐ Interview Summary (PTO-413) Paper No(s). _____
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948) 5) ☐ Notice of Informal Patent Application (PTO-152)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) 4. 6) ☐ Other: _____

DETAILED ACTION

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

2. Claims 1-20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Vu et al (U.S. Patent No. 6,393,427, hereinafter, "Vu") in view of Myers et al (U.S. Patent No. 6,374,274, hereinafter, "Myers").

With respect to claims 1, 13-14 and 18, Vu discloses the method for automatically generating a query,

defining an organized classification of document content with each class in the organized classification of document content having associated therewith a classification label, see (col. 3, lines 41-67, col. 1, lines 22-67, col. 6, lines 21-31);

each classification label corresponding to a category of information in an information retrieval system, see (col. 3, lines 41-67, col. 2, lines 2-54, col. 6, lines 21-31, Note: document classification results are saved in database based on category);

assigning the selected document content a classification label from the organized classification of content, see (col. 8, lines 16-65, col. 6, lines 21-31);

automatically formulating a query that restricts a search at the information retrieval system for information concerning the set of entities (web and database are used as entities for query) to the category of information in the information retrieval system identified by the assigned classification label, see (col. 4, lines 55-67 to col. 5, lines 1-37).

Vu does not explicitly disclose the identifying a set of entities in selected document content for searching additional information related to using the information retrieval system. However, Myers teaches that documents are identified by each entity, which teaches the entities are identified for document classification, see (col. 2, lines 28-43 and abstract). Therefore, it would have been obvious a person having ordinary skill in the art the time invention was made to include the identifying a set of entities in selected document content for searching additional information related to using the information retrieval system of Vu to have additional information that is document content entity for document classification. Because document content entity is one of the basic information for document classification or document clustering.

With respect to claims 2, Vu discloses that limiting the query by adding terms (keywords) relating to context information surrounding the set of entities in the selected document content, see (col. 5, lines 1-36).

With respect to claim 3, Vu discloses that the number of terms added is limited to a predefined number, see (col. 7, lines 9-67 to col. 8, lines 1-65).

With respect to claims 4, Vu discloses that limiting the query by adding terms defining the assigned classification label, see (FIG. 3, col. 4, lines 19-67 to col. 5, lines 1-37).

With respect to claim 5, Vu discloses that the organized classification of document content is defined using a hierarchical organization, see (FIG. 4, col. 6, lines 1-61).

With respect to claim 6, Vu discloses that using a text categorizer (classification based on keywords) to assign the classification label assigned from the organized classification of content, see (FIG. 2-4, col. 4, lines 19-67 to col. 5, lines 1-37).

With respect to claim 7, Vu discloses that extracting with the text categorizer a set of terms relating to the document content; and appending to the query ones of the set of terms extracted by the text categorizer to contextualize the query, see (FIG. 2-4, col. 4, lines 19-67 to col. 5, lines 1-37).

With respect to claim 8, Vu discloses that abbreviating the set of terms extracted by the text categorizer to a predefined limit of terms, see (FIG. 2-4, col. 4, lines 19-67 to col. 5, lines 1-37).

With respect to claim 9, Vu discloses that extracting noun phrases from the selected document content, see (col. 4, lines 55-67 to col. 5, lines 1-37);

ranking the noun phrases by those that occur most frequently in the document content, see (col. 4, lines 55-67 to col. 5, lines 1-37);

defining a subset of noun phrases by identifying those ranked noun phrases that occur more frequently than a first predefined frequency, see (col. 4, lines 55-67 to col. 5, lines 1-37); ranking those words in the subset of noun phrases by their frequency of occurrence to define an ordered list of words; defining a subset of the ordered list of words by identifying those ranked words that occur more frequently than a second predefined frequency; re-ranking the subset of words in inverse frequency to their use in the category of information in the information retrieval system identified by the assigned classification label; using only those highest ranked words in the re-ranked subset of words-to define the set of terms appended to the query, see (col. 4, lines 19-67 to col. 5, lines 1-37, Note: extracting most frequently used keywords are used to rank document classification).

With respect to claims 10-11, Vu discloses that each class in the organized classification of document content has associated therewith a characteristic vocabulary, see (col. 4, lines 19-67 to col. 5, lines 1-37).

With respect to claim 12, Vu discloses that enriching selected content of a document with personalities that identify enrichment themes, see (FIG. 9B, col. 4, lines 19-45).

With respect to claims 15-16, Vu discloses that short length aspect vector generator for generating terms relating to context information surrounding the set of entities in the selected document content; wherein the query generator adds the terms relating to the context information to limit the query, see (col. 4, lines 19-67 to col. 5, lines 1-37).

With respect to claim 17, Vu discloses that content manager for enriching the selected document content with results provided from the information retrieval system using the query, see (col. 4, lines 19-67 to col. 5, lines 1-37, Note: document modified by doing filtering and morphing, which is enriching documents).

With respect to claims 19, Vu discloses that limiting the query by adding terms (keywords) relating to context information surrounding the set of entities in the selected document content, see (col. 5, lines 1-36).

With respect to claims 20, Vu discloses that limiting the query by adding terms defining the assigned classification label, see (FIG. 3, col. 4, lines 19-67 to col. 5, lines 1-37).

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Smiley (U.S. Patent No. 5,978,811) discloses the system for data model for an information repository models data as objects, the relationships or interdependencies between the data, their physical storage or access information and rules or methods of accessing the data.

Johnson et al (U.S. Patent No. 6,212,532) discloses the system for information extraction system capable of extracting information from natural language documents. The system includes a plurality of interchangeable modules including a data preparation module for preparing a first set of raw data having class labels to be tested, the data preparation module being selected from a first type of the interchangeable modules. The system further includes a feature extraction module for extracting features from the raw data received from the data preparation module and storing the features in a vector format, the feature extraction module being selected from a second type of the interchangeable modules. A core classification module is also provided for applying a

Art Unit: 2172

learning algorithm to the stored vector format and producing therefrom a resulting classifier, the core classification module being selected from a third type of the interchangeable modules. A testing module compares the resulting classifier to a set of preassigned classes, where the testing module is selected from a fourth type of the interchangeable modules, where the testing module tests a second set of raw data having class labels received by the data preparation module to determine the degree to which the class labels of the second set of raw data approximately corresponds to the resulting classifier.

Byrd, Jr. et al (U.S. Patent No. 5,832,480) discloses the system for Descriptive canonical forms of entity types are created by scanning one or more documents in a database of a computer system to identify one or more proper names that appear in the documents as raw names. Each of the raw names has zero or more proper names, zero or more medial substrings, zero or more leading substrings, and zero or more trailing substrings. The raw names of one or more documents are "cleaned" and "split" until certain "cleaning and splitting conditions" are no longer met to obtain a list of clean and split candidate names. Anchor names are selected from the list that unambiguously represent an entity type. The anchor names have one or more entity-type attribute values. Variant names, clean and split candidate names having one or more shared attribute (values) with the anchor name, are combined with the anchor name to create an equivalence group of names that refer to the same entity. A canonical form is generated for the group from a subset of the anchor name attributes. A canonical form is created in this manner for all of the clean and split candidate names on the list.


Art Unit: 2172

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Isaac M Woo whose telephone number is (703) 305-0081. The examiner can normally be reached on 8:00-5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Kim Y Vu can be reached on (703) 305-4393. The fax phone numbers for the organization where this application or proceeding is assigned are (703) 746-7239 for regular communications and (703) 308-6606 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 305-3900.

IMW
December 13, 2002



JEAN M. CORRIELUS
PRIMARY EXAMINER